The interaction between local focusing structure and global intentions in spoken discourse

Sofia Gustafson-Capková
Department of Linguistics, Computational Linguistics
Stockholm University
S-109 61, Stockholm
Sweden
sofia@ling.su.se

ABSTRACT

The purpose of the study reported in this paper is to investigate how local focusing structure, analysed in terms of Centering Theory (Grosz, Joshi & Weinstein, 1995), and global d iscourse structure, analysed in terms of discourse segments and discourse segment purposes (Grosz & Sidner, 1986), interact. Swedish dialogue was analysed according to Centering Theory and Grosz and Sidners (1986) discourse theory. The results indicate an interaction between locally implicit elements and global intentions. Also indications concerning discourse markers varying intonation were found.

Introduction

Discourse can be described as built up from discourse building blocks called discourse segments (hereafter DS). These DS are the units for forming a hierarchical discourse structure. They are described in e.g. Grosz & Sidner (1986, hereafter G&S) where claims are made about the use of the DS in the global discourse structure as well as their connect ion to the coherence of the discourse, and in Centering Theory (Grosz, Joshi & Weinstein, 1995, hereafter CT), where claims are made about the internal structure and coherence of the DS:s.

Grosz & Sidner (1986) have applied their discourse theory to both argumentative text and task oriented dialogue, while CT traditionally has been applied to narrative text. In recent times, however, an interest for applying CT to dialogue has arisen, and some attempts to do that has been carried out (e.g. Brennan, 1998, B yron & Stent, 1998, Eckert & Strube, 1999).

In applying G&S theory, a problematic point is the importance of speaker intention, which governs both the discourse segmenting and the discourse structure. It is unclear whose perspective, that should be taken; the speakers original intention, the listeners understanding of the intention or the analysers interpretation of the intention. However, one thing that is for sure is that the analyser will certainly face a challenge if attempting to find out original speaker/listener intentions.

The major problematic issues in applying CT the issue of both utterance segmenting and discourse segmenting effects almost all other aspects of the analysis. Another problem with CT is to decide what concepts that are accessible, or *realised*, in an utterance.

The pilot study presented in this paper is of explorative character, and addresses a range of problems encountered in a combined G&S-type analysis and CT analysis of task oriented dialogue, i.e.:

- Utterance segmenting, i.e. the units between which local coherence is computed
- Discourse segmenting, i.e. larger constituents affecting the global discourse structure. These segments correlates with what Carletta et al. (1997) calls "game".
- What items that are possible centers
- The CT notion of a *realised* item

The aim of this paper is to give a picture of how those problems are connected to each other, and to outline how to refine a multiple -level analysis. It is also an attempt to apply a global and local analysis to spok en language data, and to give account for specific problems that arises by such an analysis. It is the hope of the author that results from investigations like this should help to develop e.g. instructions for more extensive investigations in the field.

1 Background

G&S and CT are two theories that give account for discourse structure and coherence, but a t different levels of the discourse. G&S mainly addresses the global discourse structure, while CT gives account for the local coherence. I will here give a short description of both theories.

G&S (1986) describe discourse as consisting of three structures : i) the linguistic structure, ii) the intentional structure and iii) the attentional state. These three structures interact, but they are still to be considered as separate structures. The interaction between them works roughly as follow: The linguistic s tructure, i.e. the string of words, is divided into discourse segments. Each segment has a Discourse Segment Purpose (DSP) which is part of the intentional structure. According to how the DSP:s are satisfied, different relations hold between the discourse segments and the attentional state is modelled out of these relations.

Thus, in G&S, disocurse segments are intentionally delimited, i.e. a discourse segment is governed by a main intention, the DSP. The range of DSP:s is unlimited. DS may be nested, an d the relations that hold between discourse segments are limited to two: i) dominance and ii) satisfaction -precedence. Dominance means in short that a discourse segment B which is part of the satisfaction of the intention governing the discourse segment A is dominated by A, i.e. A dominates B. Satisfaction -precedence on the other hand holds in the cases where the intention of a discourse segment C has to be fulfilled before the intention of the discourse segment D appears, i.e. C satisfaction-precedes D.

The relations dominance and satisfaction precedence contributes incrementally to the discourse structure and model the global coherence. This is done by stack manipulations, which could be described as modelling a temporal sequence of intentions in attentional focus in the discourse. This process will however not be closely described here.

Centering Theory is a theory, which gives account for the degree of local coherence between utterances within a discourse segment. This is made by segmenting the linguistic string into utterances and classify the transitions between them. The transitions are computed on basis of two factors: backward-looking center Cb and forward-looking center(s) Cf. Sometimes the preferred center, i.e. the highest ranked member of the Cf-list is singled out as Cp. The choice of centers is in standard CT (Grosz, Joshi & Weinstein) based on grammatical roles: subject>object>other roles. It is important to note, that the centers does not have to be explicitly present in the linguistic string (directly realized), but may also be implicitly present in the conceptual representation (realized). This means, that centers are not linguistic units, but concepts.

The four transitions are computed on basis of the C:s, as shown in Table 1.

Table 1 Table over the transitions in Centering Theory.

	$Cb(U_i) = Cb(U_{i-1})$ *or* $Cb(U_{i-1}) = [?]$	$Cb(U_i) \neq Cb(U_{i-1})$
$Cb(U_i) = Cp(U_i)$	CONTINUE	SMOOTH-SHIFT
$Cb(U_i) \neq Cp(U_i)$	RETAIN	ROUGH-SHIFT

In addition two rules are used in CT. The first is "The pronoun constraint". This rule state that if something in U_i is realized as a pronoun in U_{i+1} the Cb of U_{i+1} must also be realised with a pronoun.

The second rule states that sequences of continuation are preferred over sequences of retain. The shift transitions put generally a higer inference load upon the hearer.

The segmenting issue is certainly important also in CT, but it is not closer addressed by Grosz, Joshi & Weinstein (1995), i.e. no explicit base for the discourse segments is given here. It is however a good guess that they should be of the same nature as by Grosz & Sidner, who mentions centering as possible additional mechanism (Grosz & Sidner, 1986, p. 91).

A modified version of CT is made by Walker (1997). She has replaced the DS by something that could be described as a moving window. This means that the CT-analysis is done continually through the whole discourse, and it does not start and stop over and over again by the initiation or ending of a DS. The elements from the Cf-list are saved in a cache, which is incrementally updated in the way that new items are added and old items are erased. Walker suggests that the size of the moving window should consist of two or three sentences, or seven propositions.

Both utterance boundaries and discourse segment boundaries are difficult to delimit in spoken language. Utterances are difficult because there is often no formally correctly completed sentence structure in spontaneous speech. General strategies for segmentation of spoken discourse are prosodic phrasing, cue-words, and the use of form for referring expressions (e.g. Passonneau & Litman, 1997, Grosz & Sidner, 1986, Walker, 1997).

An additional problem in analysing dialogue is that it is not quite clear how to apply a theory like CT, mainly developed with work on narrative text, for a multi-party discourse. E.g. is the previous utterance for X the linearly previous utterance, or the previous utterance uttered by X? One has to work with at least two persons interpretations of the discourse, interpretations which do not have to be overlapping, in terms of both DS and focus of attention, i.e. one should try to keep track on whose center that is analysed.

2 Method

In order to give spoken language a discourse analysis in terms of both G&S and CT, elicited spoken dialogue was analysed. The spoken language material was from one Map Task dialogue in Swedish (Helgason). In all 60 turns from one dialogue with two speakers were analysed.

The dialogue was segmented with a pause-detecting tool, which detected silent pauses longer than 100 ms. After examination of the segmented data the analyser decided that pauses 300 ms or longer should be used as utterance boundaries. This pause length is roughly correlating with clause boundaries according to Garman (1990), who sets clause boundaries to 400 ms. The segmentation based on 300 ms or longer pauses resulted in 100 utterances. The decision was also made that change of speaker also indicated new utterance.

All linguistic units were regarded as valid, i.e. no filtering out of utterances consisting only of e.g. humming (mmm…) was done, as done by e.g. Byron & Stent (1998).

The transitions between the utterances were computed according to CT, but the ranking of centers were limited to linear appearance. When it comes to items possible carry center, 1 and 2 sg. pronouns were filtered out. After this the material was examined according to intentional content. Boundaries between DS, which correlated to certain intentions, were annotated and also the intentions were described. This resulted in 36 labelled discourse segments, which were analysed in terms of relations between different DS (dominance and satisfaction-precedence). Change of speaker was not taken to imply new DS.

3 Results and discussion

One of the main problems in discourse analysis is choosing an interpretation that is as general as possible, i.e. to try to minimise the subjectivity in the interpretation. The reason for this is the need for a possible replication of the analysis of the interpretation, i.e. the analysis should not be too bounded to the analysers subjectively based interpretation. The analyser is thus forced to keep language interpretation and discourse analysis strictly separated. This is in itself a paradox, because to analyse a stretch of discourse means to analyse an interpretation of the stretch of discourse. However, human natural language is never impersonally interpreted, it is always interpreted through the filter of a subjective human thinking, so the interpretation and the analysis blend. In fields as e.g. computational linguistics, one tries to model a pure and objective interpretation, which in fact is the most unlikely interpretation in its pureness. The question is, how is it possible to keep on searching for the most general interpretation but still avoid both subjectivity and artificiality in the interpretation, i.e. to make the claim of the analysis part of the interpretation, as objective as possible, so that it is scientifically valid, but still keep as much subjectivity as possible in the interpretation so that the outcome mimics language users as much as possible. In the subjectivity of human language understanding lies also the robustness and the generality of human language use.

To use dialogue is one way to try to delimit the degree of subjectivity in the analysis, but still allow subjectivity in the interpretation. The reason for this is that the primary task for the analyser is not to interpret the text/speech, but to understand how the current speaker interpreted what the former speaker said. This means that the analyser has a reference point for the interpretation outside herself. In following the dialogue it is also possible to follow how a person actually interprets the current speaker. The analyser is not completely alone with her own interpretation, but is able to get a glimpse of how another person interprets the utterances.

To use the pause-tool for detecting utterance boundaries was another way to try to limit the influence of subjective interpretation. The interpreter was not determining the segmentation herself, but used a kind of bootstrapping in deciding the utterance units.

The results showed that the utterance segmentation in many cases was quite good, but still, in many cases the granularity was finer than preferred. The discourse segmenting on intentional basis did not pose great problems, but perhaps that just indicates the readiness by the analyser to assign explicit intentions to certain segments. Below an overview of the segmented material is shown.

- Turns: 60
- Utterances: 100
- Discourse segments: 36

3.1 Segmenting the utterances

Cases where the pause-based utterance segmentation was not optimal could sometimes have been avoided if the intonation contour had been taken into consideration. In Example 1, given below, the speech signal was segmented at the point Utt 2. (the pause preceding that position is longer than 300 ms), but this break could have been avoided if the fact, that the intonation contour is stable (i.e. neither rising nor falling) had been taken into consideration.

Example 1

Utt. 1. .då ska vi se då har vi. .en. .en s!0 .;*karta* här framför oss. .och jag har. .;. *landstigit* på en *plats*, →
Utt. 2. .på den här ön.

In the analysis cases like Example 1 were however regarded as two utterances.

3.2 The CT analysis

After the segmentation into utterances, the 100 utterances were analysed in terms of CT. At this point in the investigation no attempt to divide the discourse into DS was made. Following Walker (1997) a continual examination of the centers and the transitions was done throughout the whole discourse.

3.2.1 Analysing the centers

Concerning the analysis of the centers the ranking based on grammatical role did not turn out to be suitable for the analysed dialogue, partly due to the fact that 1 and 2 sg. pronouns were filtered out. Instead the analyser followed three simple statements:

- All kinds of elements (e.g. complex phrases as well as single words) were ranked after the linear occurrence in the speech signal.
- Phonetic prominence was taken into account. A phonetically prominent element was given a higher rank than a phonetically non-prominent element (the pronoun constraint was however always kept).
- Coordinate and subordinate clauses were specially handled. A con-/subjunction inside an utterance started a new Cf-list, which was given higher prominence than the first list.

An example of the ranking of elements in an utterance is given in
Example 2, where the underlined "men" initiates the second Cf-list (also underlined), from where the preferred center is chosen.

Example 2

.ja. .det är ett ***aningers. .när när***mare floden än ;. .;. .kust!0. .!0kanten där men det är ***nästan mitt*** emellan.

Cb = ja<du snuddar nästan vid en flod när du är där (preceding utterance)>
Cf = 1. [närmare floden, kustkanten] 2. [mittemellan <floden & kustkanten>]
Cp = mittemellan <floden & kustkanten>

As earlier noted, Cb:s could be directly realised or realised. In Example 1 for instance some elements are present in the analysis, but not present in the utterance or in the appropriate place in the utterance. Such instances are marked out with <> in the analysis. In Example 1 there are two instances of such partly implicit elements: 1. <du snuddar nästan vid en flod när du är där> and 2. <floden & kustkanten>. In the first case, "ja" does not only seem to be a way to signal that the listener have understand, but also a way to signal that the representation of the concepts is still relevant and active, i.e. “ja” functions as a short "keep active-signal". This is found in all ja-instances. Similar findings are reported by Eckert & Strube (1999), who claim that those utterances have high relevance for grounding in dialogue. In the second case both the river (flod) and the shore (kustkanten) are introduced, but in the later part of the utterance (after "men") the focus is on the point between the both elements. However, the elements are still highly active in defining the point in between that is why the conjunction of the both concepts is analysed as present. Such partly implicit elements are frequent in the material, but also completely implicit elements. Interesting is, that in the case of a complete implicit element in an utterance as: “.och ***fortsätter*** <vägen> norrut”, the Cp in the utterance, the concept “road” (<vägen>) is a crucial concept in the formulation of the DSP (the discourse segment purposes, the intentions motivating a discourse segment). Thus, the concept could be said to be contextually highly activated, i.e. activated by the task and the situation itself, or activated on the global level. As well as we can talk about local and global focus, we can also be able to distinguish between a local and a global level of activation.

To get a view over the proportions of implicit vs. explicit reference in discourse all the Cb:s were counted and sorted as directly realised (explicitly present) or realised (implicit or partly implicit present). The result is shown in Table 2.

Table 2

	Explicitly present	partly implicitly present	fully implicitly present
N = 100	19	71	10

The above figures indicate that 81% of back reference in a discourse is implicit, which makes human communication seem like an iceberg.

The proportion of transitions between utterances was computed, and the results are given in Table 3.

Table 3 Transitions between the 100 utterances in the material

Continue	Retain	Smooth-shift	Rough-shift
47	36	10	5

Please not that one instance of Rough-shift is clearly discourse initial, so it is left out in the table above. These results will be closer discussed under the heading 3.3.

3.3 The global structure of the discourse

The global structure of the discourse was analysed in terms of the relations dominance and satisfaction-precedence between discourse segments. In making this analysis the analyser experienced a need to make a more fine-grained distinction between different instances of the relation satisfaction-precedence. Thus, the relations used were:

- Dominance, corresponds to the dominance relation between two segments (mother-daughter).
- Single pop, corresponds to two adjacent segments on the same level both without daughters (sisters without daughters). This is the same as the relation satisfaction-precedence between two sisters without daughters.
- Multiple pop, corresponds to two segments, textually adjacent but on different levels in the hierarchical analysis, i.e. the youngest daughter in one branch and a potential mother for another branch. This is the same as satisfaction-precedence between two nodes on different hierarchical levels.

As noted in Table 3 above there was an overwhelming number of the relations Continue and Retain. Both shift-transitions were quite rare, and it is worth noting that all Rough-shifts appeared either i) inside a discourse segment (4) ii) between two discourse segments related to each other by the relation dominance (1) iii) after a very clear indication that the discourse topic will change (“då så, då ska vi se”. This is the one left out in Table 3). The last alternative is possible to exclude on the basis that it is better to consider this as a new discourse and not a shift inside the same discourse. The two first alternatives however indicate, that rough-shift appears only inside a tightly defined intentional space, in the data it never appears together with a shift of the intention. It never appeared between DS:s related with Single pop or Multiple pop. A mechanism as rough-shift seems thus not to be the

appropriate way to make such a change of direction in the discourse, rather it indicates misunderstanding or a "jumping" inside one isolated intentional space.
The transitions between discourse segments is shown in Table 4 below.

Table 4 CT-Transitions at different kinds of DS boundaries.

	Continue	Retain	Smooth-shift	Rough-shift
Multiple pop	0	9	0	0
Single pop	0	6	2	1
Dominance	6	6	0	4

In investigating what could be characteristic for discourse segment boundaries with different relations all discourse segment boundaries were investigated. The results are given below.

- Multiple pop: Indicates in seven cases of nine with a combination of pause, cue-words and phonetic prominence (***och sen***/***och fortsätter***).
- Single pop: Indicates in five cases of nine with a combination of pause and cue-words (och sen / då)
- Dominance: No special preferences found.

4 Summary and further work

This investigation reported in this paper is certainly suffering from a range of weak points; for instance a larger set of data and an evaluation of inter coder reliability would be highly desirable. The analysis is now very dependent on one analysers own interpretation. The results however give quite interesting indications concerning the interaction between local focus and global intentions, e.g. the connection between the implicit centers and the intentions behind the discourse segments.
The use of pauses for utterance segmentation would certainly be better if the intonation contour could be integrated in the analysis. In the data it also seems to be a regularity in the use of intonation by the use of cue-words, i.e. the alternation between phonetically prominent and phonetically non-prominent correlates with the different relations Multiple pop and Single pop, it is however difficult to say anything for sure without analysing a larger amount of data.
Further work in this direction would, except more data, include a more thorough investigation of the ranking order. To isolate what concepts that are present, or rather accessible, in an utterance in a certain context is also indeed an important, but difficult task to attack. It would also be of interest to connect findings from analyses like this to dialogue coding, as described by e.g. Carletta et al.

5 Acknowledgments

Many thanks to Harald Berthelsen, who made the pause detecting tool, and to Petur Helgason who made the Swedish Map Task dialogues available.

6 Literature

Carletta et al. (1997): *The Reliability of a Dialogue Coding Structure Scheme*. Computational Linguistics, vol. 23 no. 1.

Berthelsen, H.: Pause detecting tool.

Brennan, S (1998): *Centering as a psychological resource for achieving joint reference in spontaneous discourse*. In Walker et al. 1998, Centering Theory in discourse.

Byron, D. & Stent, A. (1998): *A Preliminary Model of Centering in Dialog.* In the Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL'98) student session.

Eckert, M. & Strube, M (1999): *Resolving Discourse Deictic Anaphora in Dialogues*. In EACL '99.

Garman, M. (1990): *Psycholinguistics.* Cambridge University Press.

Grosz, B. & Sidner, C. (1986): *Attention, Intentions and the Structure of Discourse*. Computational Linguistics, vol 12, no 3.

Grosz, B., Joshi, A. & Weinstein, S. (1995): *Centering: A Framework for Modelling the Local Coherence of Discourse*. Computational Linguistics, vol. 21, no. 2.

Helgason, P. : *The Stockholm Corpus of Spontaneous Swedish.* Map Task corpus, Department of linguistics, Stockholm University.

Passonneau, R. & Litman, D. (1997): *Discourse Segmentation by Human and Automated Means*. Computational Linguistics, vol 23, no 1.